

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- BLURRY OR ILLEGIBLE TEXT
- SKEWED/SLATED IMAGES
- COLORED PHOTOS
- BLACK OR VERY DARK BLACK AND WHITE PHOTOS
- UNDECIPHERABLE GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PATENT ABSTRACTS OF JAPAN

(11) Publication number : 10-274996
 (43) Date of publication of application : 13.10.1998

(51) Int.CI.

G10L 3/00
 G10L 3/00
 G10L 3/00

(21) Application number : 09-081700

(22) Date of filing : 31.03.1997

(71) Applicant : TOSHIBA CORP

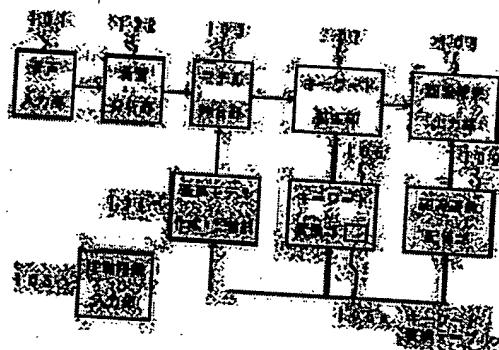
(72) Inventor : MASAI YASUYUKI
TANAKA SHINICHI

(54) VOICE RECOGNITION DEVICE

(57) Abstract:

PROBLEM TO BE SOLVED: To reduce erroneous recognition due to a storage difference, etc., in reading when a recognition device is used.

SOLUTION: This device collates a characteristic parameter series obtained by acoustic analyzing an input voice with an acoustic analytic part 102 with voice models of all key words constituting respective recognition vocabularies beforehand formed and stored in a voice model forming/storing part 104 by a model collation part 103 to recognize the input voice. In such a case, a key word expansion part 107 uses a key word conversion part 106 having a function performing the conversion between mutual key words having similarity as a sound among respective key words, and converts the key word recognized by the collation in the model collation part 103 to another key word having the similarity as the sound, and the key word recognized in the model collation part 103 is expanded, and a recognition result output part 108 decides the recognition result from the key word expanded here.



LEGAL STATUS

- [Date of request for examination] 30.10.2002
- [Date of sending the examiner's decision of rejection]
- [Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
- [Date of final disposal for application]
- [Patent number]
- [Date of registration]
- [Number of appeal against examiner's decision of rejection]
- [Date of requesting appeal against examiner's decision of rejection]
- [Date of extinction of right]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-274996

(43)公開日 平成10年(1998)10月13日

(51)Int.Cl.⁶

G 10 L 3/00

識別記号

5 6 1

5 2 1

5 7 1

F I

G 10 L 3/00

5 6 1 F

5 2 1 E

5 7 1 H

審査請求 未請求 請求項の数13 OL (全 13 頁)

(21)出願番号 特願平9-81700

(22)出願日 平成9年(1997)3月31日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 正井 康之

神奈川県川崎市幸区柳町70番地 株式会社
東芝柳町工場内

(72)発明者 田中 信一

神奈川県川崎市幸区柳町70番地 株式会社
東芝柳町工場内

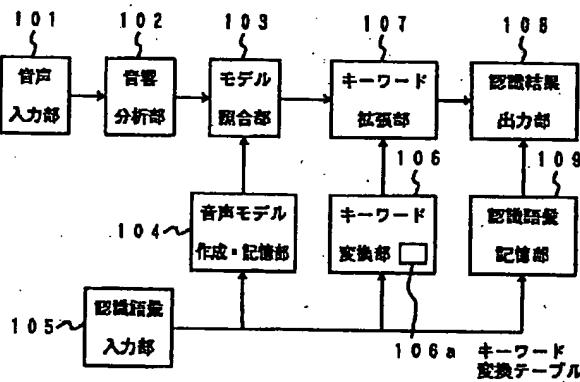
(74)代理人 弁理士 鈴江 武彦 (外6名)

(54)【発明の名称】 音声認識装置

(57)【要約】

【課題】認識装置を使用するときの読みの記憶違い等による誤認識の削減を図る。

【解決手段】入力音声を音響分析部102で音響分析して求めた特徴パラメータ系列を、音声モデル作成・記憶部104にて予め作成して記憶しておいた各認識語彙を構成する全てのキーワードの音声モデルとモデル照合部102で照合して入力音声を認識する音声認識装置において、上記各キーワードのうち音としての類似性のあるキーワード相互間の変換を行う機能を持つキーワード変換部105をキーワード拡張部107が利用して、モデル照合部102での照合で認識されたキーワードを音としての類似性のある別のキーワードに変換することで当該モデル照合部102で認識されたキーワードを拡張し、ここで拡張されたキーワードから認識結果出力部108が認識結果を決定する。



1

【特許請求の範囲】

【請求項1】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する音声認識装置において、

前記各キーワードのうち音としての類似性のあるキーワード相互間の変換を行うためのキーワード変換手段と、前記認識したキーワードを前記キーワード変換手段を用いて音としての類似性のある別のキーワードに変換することで認識したキーワードを拡張するキーワード拡張手段と、

前記拡張されたキーワードから認識結果を決定する認識結果出力手段とを具備することを特徴とする音声認識装置。

【請求項2】 前記キーワード変換手段は、前記各キーワードのうち、キーワードの読みが異なる音節数が所定数以下のキーワード同士を音としての類似性ありとして扱うことを特徴とする請求項1記載の音声認識装置。

【請求項3】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置において、

前記音声モデルとの照合により得られる認識した音声の先頭の音節と最後の音節が一致することを制約として、入力音声の最初の音節を認識する認識結果出力手段を具備することを特徴とする音声認識装置。

【請求項4】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた、認識語彙の別称の音声モデルを含む各認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置において、

登録する認識語彙の正しい読みに加えて、誤読されやすい読みを別称として登録する別称登録手段と、

前記別称登録手段により別称として登録される読みが認識語彙の読みと類似しているか否かを判定する登録語彙類似性判定手段と、

前記登録語彙類似性判定手段により類似していると判定された場合に、その旨の警告を出力する警告出力手段と、

前記別称登録手段により登録された前記別称の読みの音声モデルとの照合で入力音声が認識された場合、その読みを別称とする語彙を認識する認識結果出力手段とを具備することを特徴とする音声認識装置。

【請求項5】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する音声認識装置において、

複数のキーワードが組み合わされたキーワード列からなる音声の認識結果を複数候補出力する際に、意味的に同じキーワード別に、そのキーワード列の確からしさ順にソートして出力するキーワード別認識結果出力手段を具

備することを特徴とする音声認識装置。

【請求項6】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた、認識語彙の別称の音声モデルを含む各認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置において、登録する認識語彙の正しい読みに加えて、誤読されやすい読みを別称として登録する別称登録手段と、別称として登録してある読みを認識したときに、正しい読みを利用者に提示する認識結果出力手段を具備することを特徴とする音声認識装置。

【請求項7】 認識語彙入力手段から入力される認識語彙の読みの表記をもとに音声モデルを作成して記憶しておき、入力音声を音響分析して求めた特徴パラメータ系列を前記記憶しておいた各音声モデルと照合して、入力音声を認識する音声認識装置において、

前記認識語彙入力手段から入力される認識語彙の読みの表記を音声に変換して出力する音声出力手段を具備することを特徴とする音声認識装置。

【請求項8】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する音声認識装置に適用される音声認識方法であって、前記認識したキーワードを、前記各認識語彙を構成するキーワードの中で音としての類似性のある別のキーワードに変換することで認識したキーワードを拡張し、前記拡張したキーワードから認識結果を決定するようにしたことを特徴とする音声認識方法。

【請求項9】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置に適用される音声認識方法であって、

前記音声モデルとの照合により得られる認識した音声の先頭の音節と最後の音節が一致することを制約として、入力音声の最初の音節を認識するようにしたことを特徴とする音声認識方法。

【請求項10】 入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた、認識語彙の別称の音声モデルを含む各認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置に適用される別称登録誤り検出方法であって、

前記認識語彙の音声モデルを予め作成する際に、当該認識語彙の正しい読みに加えて、誤読されやすい読みを別称として登録し、

前記別称の読みの登録時には、当該別称の読みが、既に登録済みの認識語彙の読みと類似しているか否かを判定し、

類似していると判定した場合には、別称登録誤りを示す警告を出力するようにしたことを特徴とする別称登録誤り検出方法。

【請求項11】 入力音声を音響分析して求めた特徴パ

3

ラメータ系列を予め作成しておいた認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置における認識結果出力方法であって、

複数のキーワードが組み合わされたキーワード列からなる音声の認識結果を複数候補出力する際に、意味的に同じキーワード別に、そのキーワード列の確からしさ順にソートして出力するようにしたことを特徴とする認識結果出力方法。

【請求項12】入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた、認識語彙の別称の音声モデルを含む各認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置に適用される認識語彙の読みの提示方法であって、

別称として登録してある認識語彙の読みを認識したときには、当該認識語彙の正しい読みを利用者に提示するようにしたことを特徴とする認識語彙の読みの提示方法。

【請求項13】認識語彙入力手段から入力される認識語彙の読みの表記をもとに音声モデルを作成して記憶しておき、入力音声を音響分析して求めた特徴パラメータ系列を前記記憶しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する音声認識装置に適用される認識語彙の読みの提示方法であつて、

前記認識語彙入力手段から認識語彙の読みの表記が入力された際に、当該表記を音声に変換して出力するようにしたことを特徴とする認識語彙の読みの提示方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する音声認識装置に係り、特に認識語彙の読みの登録間違いや、装置使用時の読みの記憶違いによる誤認識を削減するのに好適な音声認識装置に関する。

【0002】

【従来の技術】一般に、入力音声を認識する音声認識装置では、当該装置での認識対象となる語彙(認識語彙)の読みを入力することで、その読みから、対応する認識語彙を構成するキーワードの音声モデルを予め作成し、入力音声の認識のため記憶しておくようになっている。この種の音声認識装置での入力音声の認識は、次のように行われる。

【0003】まず入力音声を音響分析して特徴パラメータ系列を求める。次に、求めた入力音声の特徴パラメータ系列を予め作成しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する。

【0004】このような音声認識装置においては、従来は、認識語彙の読みを誤って登録した場合には、使用時に正しい読みを発声しても正しく認識できないという問

10

20

30

40

50

題があった。また、認識語彙の登録時には正しい読みを登録しておいても、使用時に誤った読みを発声すると正しく認識されないという問題もあった。

【0005】

【発明が解決しようとする課題】上記したように従来の音声認識装置では、認識語彙の読みを誤って登録すると、使用時に正しい読みを発声しても正しく認識できず、逆に認識語彙の登録時に正しい読みを登録しておいても、使用時に誤った読みを発声すると正しく認識されないという問題があった。

【0006】本発明は上記事情を考慮してなされたものでその目的は、認識語彙の読みの登録間違いや、装置使用時の読みの記憶違い等に起因する認識性能の低下を防ぐことができる音声認識装置を提供することにある。

【0007】本発明の他の目的は、利用者が、認識語彙の読みを誤って登録することを未然に防止でき、また、使用時に誤った読みで発声しても正しく認識できる音声認識装置を提供することにある。本発明の更に他の目的は、利用者による認識候補の選択が誤りなく効率的に行える音声認識装置を提供することにある。

【0008】

【課題を解決するための手段】本発明の第1の観点に係る音声認識装置は、入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する音声認識装置において、上記各キーワードのうち音としての類似性のあるキーワード相互間の変換を行うためのキーワード変換手段と、上記認識したキーワードを上記キーワード変換手段を用いて音としての類似性のある別のキーワードに変換することで認識したキーワードを拡張するキーワード拡張手段と、上記拡張されたキーワードから認識結果を決定する認識結果出力手段とを備えたことを特徴とする。ここで、音としての類似性があることの判定条件として、上記各キーワードのうち、キーワードの読みが異なる音節数が所定数以下、例えば1音節以下のキーワード同士であることを適用するとよい。

【0009】このような構成においては、キーワードを音としての類似性に着目して拡張することにより、キーワードの認識誤りや話者の言い間違いによる認識性能の低下を効果的に防ぐことが可能となる。

【0010】本発明の第2の観点に係る音声認識装置は、入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置において、上記音声モデルとの照合により得られる認識した音声の先頭の音節と最後の音節が一致することを制約として、入力音声の最初の音節を認識する認識結果出力手段を備えたことを特徴とする。

【0011】このような構成においては、入力音声の先

5

頭の音節と最後の音節が一致するという制約のもとで入力音声の先頭の音節を認識することにより、非常に精度の高い音節認識を実現できる。また、各音節を入力するときに発声する言葉を覚える必要がないので、誰でもすぐに使用することができる。

【0012】本発明の第3の観点に係る音声認識装置は、入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた、認識語彙の別称の音声モデルを含む各認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置において、登録する認識語彙の正しい読みに加えて、誤読されやすい読みを別称として登録する別称登録手段と、別称として登録される読みが認識語彙の読みと類似しているか否かを判定する登録語彙類似性判定手段と、この登録語彙類似性判定手段により類似していると判定された場合に、その旨の警告を出力する警告出力手段と、上記別称登録手段により登録された別称の読みの音声モデルとの照合で入力音声が認識された場合、その読みを別称とする語彙を認識する認識結果出力手段とを備えたことを特徴とする。

【0013】このような構成においては、認識語彙入力手段から入力して登録される認識語彙の別称を別称登録手段により登録する際に、別称が他の語彙と類似していないかを登録語彙類似性判定手段にて判断し、警告出力手段から利用者に知らせることにより、別称登録による認識性能の低下を未然に防ぐことができる。このとき、該当する別称登録を中止させるとよい。

【0014】本発明の第4の観点に係る音声認識装置は、入力音声を音響分析して求めた特徴パラメータ系列を予め作成しておいた各認識語彙を構成するキーワードの音声モデルと照合して、入力音声を認識する音声認識装置において、複数のキーワードが組み合わされたキーワード列からなる音声の認識結果を複数候補出力する際に、意味的に同じキーワード別に、そのキーワードの確からしさ順にソートして出力するキーワード別認識結果出力手段を備えたことを特徴とする。

【0015】このような構成においては、キーワード別に複数の認識結果を類似度の大きい順(あるいは距離の小さい順)に出力することにより、候補選択を効率よく行うことができる。

【0016】本発明の第5の観点に係る音声認識装置は、入力音声を音響分析して求めた特徴パラメータ系列と予め作成しておいた、認識語彙の別称の音声モデルを含む各認識語彙の音声モデルと照合して、入力音声を認識する音声認識装置において、登録する認識語彙の正しい読みに加えて、誤読されやすい読みを別称として登録する別称登録手段と、別称として登録してある読みを認識したときに、正しい読みを利用者に提示する認識結果出力手段を備えたことを特徴とする。

【0017】このような構成においては、別称での読みの音声モデルとの照合に基づく認識結果出力時に、正し

6

い読みを出力することにより、話者が別称で覚えていた言葉の正しい読みを当該話者に覚えさせることができ、以後正しい読みで入力できるようになる。

【0018】本発明の第6の観点に係る音声認識装置は、認識語彙入力手段から入力される認識語彙の読みの表記をもとに音声モデルを作成して記憶しておき、入力音声を音響分析して求めた特徴パラメータ系列を上記記憶しておいた各音声モデルと照合して、入力音声を認識する音声認識装置において、上記認識語彙入力手段から入力される認識語彙の読みの表記を音声に変換して出力する音声出力手段を備えたことを特徴とする。

【0019】このような構成においては、認識語彙入力手段から認識語彙として例えば「竹芝」の読みを登録するときに、誤って「たけしば」と入力すると、音声出力手段により「たけしば」と音声で出力されるので、話者(認識語彙登録者)は読みの入力間違いに容易に気づくことができ、読みの入力誤りによる認識性能の低下を未然に防ぐことができる。

【0020】

【発明の実施の形態】以下、本発明の実施の形態につき図面を参照して説明する。

【第1の実施形態】図1は本発明の第1の実施形態に係る音声認識装置の概略構成を示すブロック図である。

【0021】図1の音声認識装置において、音声入力部101から入力された音声は、音響分析部102で特徴パラメータに変換される。音声認識に使用される代表的な特徴パラメータとしては、バンドパスフィルタやフーリエ変換によって求めることができるパワースペクトルや、LPC(線形予測)分析によって求めたケプストラム係数などがよく用いられるが、ここではその特徴パラメータの種類は問わない。

【0022】音響分析部102は求めた入力音声の特徴パラメータ系列をモデル照合部103に出力する。モデル照合部103は、認識語彙入力部105から入力された認識語彙に従って予め作成して音声モデル作成・記憶部104に記憶しておいた認識対象とするキーワード(認識語彙を構成するキーワード)の各音声モデルと上記入力音声の特徴パラメータ系列との類似度あるいは距離を求める演算を行う。

【0023】モデル照合部103の照合方法としては、音声モデルも特徴パラメータ系列で表現しておき、DP(動的計画)法で音声モデルの特徴パラメータ系列と入力音声の特徴パラメータ系列の距離を求める手法や、HMM(隠れマルコフモデル)を用いて音声モデルを表現しておき、入力音声の特徴パラメータ系列が入力されたときの各音声モデルの確率を計算する手法などが広く使用されているが、特に手法は問わない。

【0024】認識語彙入力部105は、認識語彙と、認識語彙を構成する各キーワードの音声モデルを作成するために必要な、認識語彙の各キーワードへの分割情報

7

(キーワード分割情報) と、各キーワードの読み情報を入力するためのものであり、キーボードやファイルなどで実現することができる。認識語彙入力部105から入力された認識語彙は認識語彙記憶部109に登録される。

【 0025 】 キーワード 変換部106は、認識語彙入力部105から入力されたキーワード分割情報から各キーワードを抽出し、キーワード間の音の類似性と品詞などの属性に基づいて選択される、例えば音の類似性があつて且つ品詞が同じキーワードの変換テーブル(キーワード変換テーブル) 106aを作成し、記憶しておくためのものである。

【 0026 】 キーワード 拡張部107は、モデル照合部103で得られた(類似度あるいは距離付き) キーワードを、キーワード 変換部106によりキーワード 変換テーブル106aに従って音の類似性のある他のキーワードに変換させ、キーワードの拡張を行う。

【 0027 】 認識結果出力部108は、モデル照合部103で求めた各音声モデルとキーワード 拡張部107で拡張して得られたキーワードが組み合わされたキーワード列に対する類似度(あるいは距離) をある条件(例えば類似度の大きさ) のもとでソーティングして、認識語彙記憶部109に記憶されている認識語彙の中で、類似度が最大(あるいは距離が最小) となる認識対象のカテゴリを認識結果として出力する。なお、上記ソーティングの制約として、例えば人の氏名は、会社名より優先されるなどを適用してもよい。

【 0028 】 以上に述べた図1の構成の音声認識装置の具体的動作を、当該音声認識装置で認識対象とする語彙、即ち認識語彙が、「 佐藤商店」、「 加藤食堂」、「 田中書店」の3種類である場合を例にとり説明する。

【 0029 】 この場合、認識語彙入力部105から上記3種類の認識語彙が入力されることになるが、本実施形態では、その認識語彙を構成する各キーワードの音声モデルが(音声モデル作成・記憶部104にて) 作成可能なように、「 佐藤- 商店」、「 加藤- 食堂」、「 田中- 書店」のように、認識語彙中にキーワード分割記号(キーワード分割情報) 「 - 」が挿入されて入力される。

【 0030 】 音声モデル作成・記憶部104は、認識語彙入力部105からキーワード分割記号「 - 」が挿入された認識語彙「 佐藤- 商店」、「 加藤- 食堂」、「 田中- 書店」が入力されると、各認識語彙について、その語彙中に挿入されたキーワード分割記号「 - 」に従って、その語彙を構成するキーワードに分割する。ここでは、上記3種類の認識語彙が、「 佐藤」、「 加藤」、「 田中」、「 商店」、「 食堂」、「 書店」の6つのキーワードに分割される。音声モデル作成・記憶部104は、この6つのキーワード「 佐藤」、「 加藤」、「 田中」、「 商店」、「 食堂」、「 書店」について、それぞれ音声モデルを作成し、記憶する。

10

20

30

40

50

8

【 0031 】 これと同時に、キーワード 変換部106は、認識語彙入力部105から入力されたキーワード分割記号付きの認識語彙「 佐藤- 商店」、「 加藤- 食堂」、「 田中- 書店」から得られる上記6つのキーワード「 佐藤」、「 加藤」、「 田中」、「 商店」、「 食堂」、「 書店」について音声の類似性を調べて、類似性のあるキーワードを抽出し、キーワード 変換テーブル106aを作成する。ここでは、キーワード 変換テーブル106aの作成規則を、キーワードの読みが異なる音節数が所定数以下、例えば1音節以下のキーワード同士を音声の類似性ありとして、当該テーブル106aに登録するものとする。この場合、「 佐藤」と「 加藤」、「 商店」と「 書店」が類似性ありと抽出され、図2に示すようなキーワード 変換テーブル106aが作成される。

【 0032 】 すると、音声認識時に、例えば「 佐藤商店」と入力された場合に、モデル照合部103での照合結果が「 佐藤」と「 書店」であったとすると、キーワード 拡張部107では、「 佐藤」と「 書店」について、キーワード 変換部106によりキーワード 変換テーブル106aに従り「 佐藤→加藤」、「 書店→商店」のキーワード 変換を行わせ、モデル照合部103での照合結果として「 佐藤」と「 書店」の他に、「 加藤」と「 商店」もあるかのように、キーワードの拡張を行う。

【 0033 】 キーワード 拡張部107により拡張されたキーワードの組み合わせの中には、認識語彙記憶部109に記憶されている認識語彙と一致するものとして、「 佐藤商店」がある。したがって、モデル照合部103での照合結果が「 佐藤」と「 書店」であったにも拘らず、認識結果出力部108では、「 佐藤商店」を正しく認識して出力することができる。

【 0034 】 これに対し、キーワード 変換部106とキーワード 拡張部107がなく、キーワードの拡張が行われない場合には、モデル照合部103での照合結果である「 佐藤」と「 書店」で構成される「 佐藤書店」は認識語彙記憶部109には存在しないので、「 佐藤商店」を正しく認識することはできない。

【 0035 】 なお、キーワード 変換により得られたキーワードの音声モデルとの照合では、類似度を一定値あるいは一定割合低くするとよい。以上は、話者が「 佐藤商店」と発声したのに対して、モデル照合部103で「 佐藤」「 書店」と誤った照合結果が得られた場合でも、音の類似性に着目したキーワードの拡張により「 佐藤商店」を正しく認識できる例について述べた。本実施形態では、同様にして、話者が「 佐藤商店」を「 佐藤書店」と言い間違った場合にも、音の類似性に着目したキーワードの拡張により「 佐藤商店」を正しく認識することができる。

【 0036 】 このように本実施形態においては、キーワードを音としての類似性に着目して拡張することにより、キーワードの認識誤りや話者の言い間違いによる認

識性能の低下を効果的に防ぐことができる。

【第2の実施形態】図3は本発明の第2の実施形態に係る音声認識装置の概略構成を示すブロック図である。

【0037】図3の音声認識装置において、音声入力部201から入力された音声は、音響分析部202で特徴パラメータに変換される。音声認識に使用される代表的な特徴パラメータとしては、バンドパスフィルタやフーリエ変換によって求めることができるパワースペクトルや、LPC(線形予測)分析によって求めたケプストラム係数などがよく用いられるが、ここではその特徴パラメータの種類は問わない。

【0038】音響分析部202は求めた入力音声の特徴パラメータ系列をモデル照合部203に出力する。モデル照合部203は、音声モデル記憶部204に記憶されている全ての音節の任意の長さの音節列の音声モデルと特徴パラメータ系列の類似度あるいは距離を求める演算を行う。

【0039】モデル照合部203の照合方法としては、音声モデルも特徴パラメータ系列で表現しておき、DP(動的計画)法で音声モデルの特徴パラメータ系列と入力音声の特徴パラメータ系列の距離を求める手法や、HMM(隠れマルコフモデル)を用いて音声モデルを表現しておき、入力音声の特徴パラメータ系列が入力されたときの各音声モデルの確率を計算する手法などが広く使用されているが、特に手法は問わない。

【0040】認識結果出力部205は、モデル照合部203での照合結果をもとに、制約条件記憶部206に記憶されている制約条件に従って、例えば先頭の音節と最後の音節が一致する音節列について、類似度(あるいは距離)をある条件のもとでソーティングして、類似度が最大(あるいは距離が最小)となる音節列の先頭の音節を認識結果として出力する。

【0041】以上に述べた図3の構成の音声認識装置の具体的動作を、例えば、「あさひ(朝日)のあ」と発声した場合を例にとり説明する。まず、話者が「あさひ(朝日)のあ」と発声した結果、モデル照合部203にて図4に示すような音節列と類似度、即ち類似度が8.6の音節列「あ」「さ」「ひ」「の」「あ」と、類似度が9.2の音節列「う」「さ」「ひ」「の」「あ」とが得られたとする。

【0042】この場合、入力音声の先頭の音節と最後の音節が一致するという制約を設けないで、認識結果出力部205から類似度が最大となる音節列の先頭の音節を認識結果として出力するならば、入力音声の先頭の音節とは異なる誤った音節「う」が出力されることになる。

【0043】これに対して本実施形態では、制約条件記憶部206に記憶されている制約条件により、先頭と最後の音節が一致するという制約を設けてあるため、認識結果出力部205での認識結果は音節「あ」となり、入力音声の先頭の音節を正しく認識することができる。し

かも、先頭と最後の音節が一致するという制約のもとで、入力音声の最初の音節を認識することから、この例のように音節「あ」を入力するときに発声する音声は、「朝日のあ」だけではなく、「あひるのあ」、更には「あじあ(アジア)」など、単に先頭の音節と最後の音節が同じであればよい。

【0044】このように本実施形態においては、入力音声の先頭の音節と最後の音節が一致するという制約のもとで入力音声の先頭の音節を認識することにより、非常に精度の高い音節認識を実現できる。また、各音節を入力するときに発声する言葉を覚える必要がないので、誰でもすぐに使用することができる。

【第3の実施形態】図5は本発明の第3の実施形態に係る音声認識装置の概略構成を示すブロック図である。

【0045】図5の音声認識装置において、音声入力部301から入力された音声は、音響分析部302で特徴パラメータに変換される。音声認識に使用される代表的な特徴パラメータとしては、バンドパスフィルタやフーリエ変換によって求めることができるパワースペクトルや、LPC(線形予測)分析によって求めたケプストラム係数などがよく用いられるが、ここではその特徴パラメータの種類は問わない。

【0046】音響分析部302は求めた入力音声の特徴パラメータ系列をモデル照合部303に出力する。モデル照合部303は、認識語彙入力部305から入力された認識語彙に従って予め作成して音声モデル作成・記憶部304に記憶しておいた認識対象とするキーワードの各音声モデルと上記入力音声の特徴パラメータ系列との類似度あるいは距離を求める演算を行う。

【0047】モデル照合部303の照合方法としては、音声モデルも特徴パラメータ系列で表現しておき、DP(動的計画)法で音声モデルの特徴パラメータ系列と入力音声の特徴パラメータ系列の距離を求める手法や、HMM(隠れマルコフモデル)を用いて音声モデルを表現しておき、入力音声の特徴パラメータ系列が入力されたときの各音声モデルの確率を計算する手法などが広く使用されているが、特に手法は問わない。

【0048】認識結果出力部308は、モデル照合部303で求めた各認識語彙に対する類似度が最大(あるいは距離が最小)となる語彙を認識結果として出力する。認識語彙入力部305は、認識したい語彙とその読みを登録するためのものであり、キーボードやファイルなどで実現することができる。

【0049】一方、例えば登録したい地名として、認識語彙入力部305から「神戸(かんべ)」を登録する際に、「神戸」の読みとしては「こうべ」の方が一般的であり、「こうべ」と誤読される可能性が高いと判断した場合には、「神戸(かんべ)」の誤読されやすい読み、即ち別称として「こうべ」を別称登録部306から登録する。この別称登録部306を、例えばキーボードで構

成して、利用者からの当該キーボードの操作により別称を登録(入力)するようにするしてもよいし、語彙からその読みを検索することができるテーブルを予め作成して別称登録部306に設けておき、複数の読みが存在する場合には、別称をそのテーブルから別称登録部306内部で自動生成(入力)する構成としてもよい。

【0050】登録語彙類似性判定部307は、認識語彙入力部305から登録された全ての語彙の読みと別称登録部306での別称登録により登録される読みとの類似性を判断する。もし、別称登録される読みとの類似性のある(読みが登録された)語彙が存在する場合には、登録語彙類似性判定部307は警告出力部309により利用者に警告したり、別称の登録の中止を行う。この登録語彙類似性判定部307での読みの類似性の判定には、例えば読みの音節の相違が1音節以下などの条件が適用可能である。

【0051】このように本実施形態においては、認識語彙入力部305から入力して登録される認識語彙の別称を別称登録部306により登録する際に、別称が他の語彙と類似していないかを登録語彙類似性判定部307にて判断し、警告出力部309から利用者に知らせることにより、別称登録による認識性能の低下を未然に防ぐことができる。例えば、上記した「神戸(かんべ)」ではなくて、一般的な神戸(こうべ)」が認識語彙として登録されているにも拘らず、「神戸」の別称として「こうべ」を登録した場合には、「神戸(かんべ)」と「神戸(こうべ)」の識別はできなくなるが、図5の音声認識装置では、このような問題を回避することができる。

【0052】なお、別称登録部306により登録される別称に類似の認識語彙がない場合、例えば認識語彙「神戸(かんべ)」の別称として「こうべ」を登録する場合には、登録語彙類似性判定部307にて類似語彙がないものと判断されて別称登録が許可され、音声モデル作成・記憶部304には、認識語彙「神戸(かんべ)」の音声モデルとは別に、認識語彙「神戸(かんべ)」の別称「こうべ」の音声モデルが記憶される。この場合、「神戸(かんべ)」を誤って「こうべ」と発声しても、モデル照合部303で(音声モデル作成・記憶部304内の)「神戸(かんべ)」の別称の「こうべ」(の音声モデル)と照合されることで、「神戸(かんべ)」が認識される。

【第4の実施形態】図6は本発明の第4の実施形態に係る音声認識装置の概略構成を示すブロック図である。

【0053】図6の音声認識装置において、音声入力部401から入力された音声は、音響分析部402で特徴パラメータに変換される。音声認識に使用される代表的な特徴パラメータとしては、バンドパスフィルタやフーリエ変換によって求めることができるパワースペクトルや、LPC(線形予測)分析によって求めたケプストラム係数などがよく用いられるが、ここではその特徴パラ

メータの種類は問わない。

【0054】音響分析部402は求めた入力音声の特徴パラメータ系列をモデル照合部403に出力する。モデル照合部403は、認識語彙入力部405から入力された認識語彙に従って予め作成して音声モデル作成・記憶部404に記憶しておいた認識対象とするキーワード

(認識語彙を構成するキーワード)の各音声モデルと上記入力音声の特徴パラメータ系列との類似度あるいは距離を求める演算を行う。

【0055】モデル照合部403の照合方法としては、音声モデルも特徴パラメータ系列で表現しておき、DP(動的計画)法で音声モデルの特徴パラメータ系列と入力音声の特徴パラメータ系列の距離を求める手法や、HMM(隠れマルコフモデル)を用いて音声モデルを表現しておき、入力音声の特徴パラメータ系列が入力されたときの各音声モデルの確率を計算する手法などが広く使用されているが、特に手法は問わない。

【0056】キーワード別認識結果出力部406は、モデル照合部403で求めた各キーワードに対する類似度(あるいは距離)に従い、認識語彙入力部405から入力されて認識語彙記憶部407に記憶されている語彙を意味的に同じキーワード別にソーティングし、類似度が最大(あるいは距離が最小)となる複数の語彙を認識結果として出力する。

【0057】例えば、認識語彙記憶部407内に、認識語彙として「田中ホテル」、「佐藤ホテル」、「加藤ホテル」、「田中酒店」、「佐藤酒店」、「田中ガソリンスタンド」の6種類が登録されている場合に、音声認識するキーワードとして、「田中」「佐藤」「加藤」「ホテル」「酒店」「ガソリンスタンド」の6つキーワードを考える。

【0058】ここで、もし「田中ホテル」と発声された場合に、モデル照合部403にて得られる認識結果と類似度が図7に示すようになったものとする。この場合、キーワード別認識結果出力部406が、図8(b)に示すように、単純にキーワードが組み合わされた(認識語彙記憶部407に記憶されている語彙に一致する)キーワード列に対する類似度の和の大きい順に複数の候補を出力したのでは(従来の出力方式)、「ホテル」や「酒店」が混在しているため候補選択時にわかりにくいという問題がある。

【0059】これに対して本実施形態では、キーワード別認識結果出力部406は、例えば業種を表すキーワードの類似度が予め定められた閾値以上となるキーワード列を、当該業種を表すキーワード別に出力する。例えば、類似度が100以上の業種を表すキーワード別(ここでは「ホテル」と「酒店」の各キーワード別)に表示すると、図8(b)のように表示することができ、視認性良く候補を表示することができる。

【0060】このように本実施形態においては、キーワ

13

ード別に複数の認識結果を類似度の大きい順(あるいは距離の小さい順)に出力することにより、候補選択を効率よく行うことができる。

【第5の実施形態】図9は本発明の第5の実施形態に係る音声認識装置の概略構成を示すブロック図である。

【0061】図9の音声認識装置において、音声入力部501から入力された音声は、音響分析部502で特徴パラメータに変換される。音声認識に使用される代表的な特徴パラメータとしては、バンドパスフィルタやフーリエ変換によって求めることができるパワースペクトルや、LPC(線形予測)分析によって求めたケプストラム係数などがよく用いられるが、ここではその特徴パラメータの種類は問わない。

【0062】音響分析部502は求めた入力音声の特徴パラメータ系列をモデル照合部503に出力する。モデル照合部503は、認識語彙入力部505から入力された認識語彙に従って予め作成して音声モデル作成・記憶部504に記憶しておいた認識対象とするキーワードの各音声モデル(ここでは、認識語彙の別称の音声モデルを含む各認識語彙の音声モデル)と上記入力音声の特徴パラメータ系列の類似度あるいは距離を求める演算を行う。

【0063】モデル照合部503の照合方法としては、音声モデルも特徴パラメータ系列で表現しておき、DP(動的計画)法で音声モデルの特徴パラメータ系列と入力音声の特徴パラメータ系列の距離を求める手法や、HMM(隠れマルコフモデル)を用いて音声モデルを表現しておき、入力音声の特徴パラメータ系列が入力されたときの各音声モデルの確率を計算する手法などが広く使用されているが、特に手法は問わない。

【0064】認識語彙入力部505は、認識したい語彙とその読みを登録するためのものであり、キーボードやファイルなどで実現することができる。一方、例えば登録したい地名として、認識語彙入力部505から「神戸(かんべ)」を登録する際に、「神戸」の読みとしては「こうべ」の方が一般的であり、「こうべ」と誤読される可能性が高いと判断した場合には、「神戸(かんべ)」の誤読されやすい読み、即ち別称として「こうべ」を別称登録部506から登録する。この別称登録部506を、例えばキーボードで構成して、利用者からの当該キーボードの操作により別称を登録(入力)するようにするしてもよいし、語彙からその読みを検索することができるテーブルを予め作成して別称登録部506に設けておき、複数の読みが存在する場合には、別称をそのテーブルから別称登録部506内部で自動生成(入力)する構成としてもよい。

【0065】別称登録部506から認識語彙「神戸(かんべ)」の別称として「こうべ」を登録すると、音声モデル作成・記憶部504には、認識語彙「神戸(かんべ)」の音声モデルとは別に、認識語彙「神戸(かん

10

20

30

40

50

14

べ)」の別称「こうべ」の音声モデルが記憶される。ここで、「こうべ」の音声モデルには、「神戸(かんべ)」の別称であることを示すフラグ情報が付される。

【0066】そこで、「神戸(かんべ)」を誤って「こうべ」と発声しても、モデル照合部503で(音声モデル作成・記憶部504内の)「神戸(かんべ)」の別称の「こうべ」の音声モデルと照合されることで、「神戸(かんべ)」が認識される。ここで、「こうべ」の音声モデルには、上記したように「神戸(かんべ)」の別称であることを示すフラグ情報が付加されており、モデル照合部503で「こうべ」の音声モデルとの照合が行われた場合、その照合結果には当該フラグ情報が付されて認識結果出力部507に渡される。これにより認識結果出力部507は、モデル照合部503で認識されたキーワードは正しい読みでなくて別称であることを識別し、認識結果「神戸」に正しい読み「かんべ」を附加して、表示または音声で出力する。

【0067】このように本実施形態においては、認識結果出力時に、正しい読みを出力することにより、話者が別称で覚えていた言葉の正しい読みを当該話者に覚えさせることができ、以後正しい読みで入力できるようになる。

【第6の実施形態】図10は本発明の第6の実施形態に係る音声認識装置の概略構成を示すブロック図である。

【0068】図10の音声認識装置において、音声入力部601から入力された音声は、音響分析部602で特徴パラメータに変換される。音声認識に使用される代表的な特徴パラメータとしては、バンドパスフィルタやフーリエ変換によって求めることができるパワースペクトルや、LPC(線形予測)分析によって求めたケプストラム係数などがよく用いられるが、ここではその特徴パラメータの種類は問わない。

【0069】音響分析部602は求めた入力音声の特徴パラメータ系列をモデル照合部603に出力する。モデル照合部603は、認識語彙入力部605から入力された認識語彙の読み(仮名、カタカナ、あるいはローマ字などの表記で入力される認識語彙の読み)に従って予め作成して音声モデル作成・記憶部604に記憶しておいた認識対象とするキーワード(認識語彙を構成するキーワード)の各音声モデルと上記入力音声の特徴パラメータ系列との類似度あるいは距離を求める演算を行う。

【0070】モデル照合部603の照合方法としては、音声モデルも特徴パラメータ系列で表現しておき、DP(動的計画)法で音声モデルの特徴パラメータ系列と入力音声の特徴パラメータ系列の距離を求める手法や、HMM(隠れマルコフモデル)を用いて音声モデルを表現しておき、入力音声の特徴パラメータ系列が入力されたときの各音声モデルの確率を計算する手法などが広く使用されているが、特に手法は問わない。

【0071】音声出力部607は、認識語彙入力部60

15

5から認識語彙の読みが入力された際に、その読みを音声に変換して出力する。音声出力部607による音声出力は、例えば日本語の全ての音節について音声を記憶しておく、上記入力された読みに従って、記憶された音声を接続することにより実現することができる。

【0072】例えば、認識語彙「竹芝」の読みとして認識語彙入力部605から誤って「たけしば」と入力したとすると、その誤った読み「たけしば」の音声モデルが音声モデル作成・記憶部604で作成・記憶されるため、認識時に「たけしば」と発声しても正しく認識できなくなる。

【0073】これに対して本実施形態によれば、認識語彙入力部605から認識語彙「竹芝」の読みを登録するときに、誤って「たけしば」と入力すると、音声出力部607により「たけしば」と音声で出力してくれるので、話者(認識語彙登録者)は読みの入力間違いに容易に気づくことができ、読みの入力誤りによる認識性能の低下を未然に防ぐことができる。

【0074】以上に述べた図1、図3、図5、図6、図9、図10の構成の音声認識装置の各部の機能は、コンピュータ、例えば内蔵型マイクロホンが組み込まれた、あるいはマイクロホン入力端子が設けられた音声入力機能を持つパソコン用コンピュータを、上記音声認識装置が持つ各処理部として機能させるためのプログラムを記録した、CD-ROM、フロッピーディスク、メモリカード等の記録媒体を用い、当該記録媒体をパソコン用コンピュータに装着して、当該記録媒体に記録されているプログラムをパソコン用コンピュータで読み取り実行させることにより実現される。また、上記プログラムは、記録媒体に限らず、例えば通信回線からダウンロードされるものであっても構わない。

【0075】

【発明の効果】以上詳述したように本発明によれば、キーワードを音としての類似性に着目して拡張するようにしたので、キーワードの認識誤りや話者の言い間違いによる認識性能の低下を効果的に防ぐことができる。

【0076】また、本発明によれば、入力音声の先頭の音節と最後の音節が一致するという制約のもとで入力音声の先頭の音節を認識するようにしたので、非常に精度の高い音節認識を実現できる。また、各音節を入力するときに発声する言葉を覚える必要がないので、誰でもすぐに誤りなく使用することができる。

【0077】また、本発明によれば、認識語彙の別称を登録する際に、その別称が他の登録語彙と類似していないかを調べて、類似している場合には利用者に知らせることにより、別称登録による認識性能の低下を未然に防ぐことができる。

【0078】また、本発明によれば、キーワード別に複数の認識結果を一定の条件でソーティングして出力するようにしたので、候補選択を効率よく行うことができ、

10

16

候補選択誤り等の発生を防ぐことができる。

【0079】また、本発明によれば、認識結果出力時に、正しい読みを出力するようにしたので、話者が別称で覚えていた言葉の正しい読みを当該話者に覚えさせることができ、以後正しい読みで入力できるようになる。

【0080】また、本発明によれば、認識語彙の登録のために、その読みの表記を入力する際には、その表記を音声に変換して出力するようにしたので、話者は読みの入力間違いがあった場合には、その読みを音声で出力するようにしたので、話者は、読みの表記の入力間違いがあつても、その間違いを音声出力内容から直ちに気づくことができ、読みの入力誤りによる認識性能の低下を未然に防ぐことができる。このように本発明においては、認識語彙の読みの登録間違いや、装置使用時の読みの記憶間違い等に起因する認識性能の低下を防ぐことができる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態に係る音声認識装置の概略構成を示すブロック図。

20

【図2】図1中のキーワード変換テーブル106aの内容例を示す図。

【図3】本発明の第2の実施形態に係る音声認識装置の概略構成を示すブロック図。

【図4】「あさひのあ」と発声された場合に図3中のモデル照合部203で得られる音節列と類似度の一例を示す図。

30

【図5】本発明の第3の実施形態に係る音声認識装置の概略構成を示すブロック図。

【図6】本発明の第4の実施形態に係る音声認識装置の概略構成を示すブロック図。

【図7】「田中ホテル」と発声された場合に図6中のモデル照合部403にて得られる各キーワードと類似度の一例を示す図。

40

【図8】図7の認識結果と類似度に基づく認識結果表示例を従来方式の認識結果表示例と対比させて示す図。

【図9】本発明の第5の実施形態に係る音声認識装置の概略構成を示すブロック図。

【図10】本発明の第6の実施形態に係る音声認識装置の概略構成を示すブロック図。

【符号の説明】

101, 201, 301, 401, 501, 601…音声入力部

102, 202, 302, 402, 502, 602…音響分析部

103, 203, 303, 403, 503, 603…モデル照合部

104, 304, 404, 504, 604…音声モデル作成・記憶部

105, 305, 405, 505, 605…認識語彙入力部

50

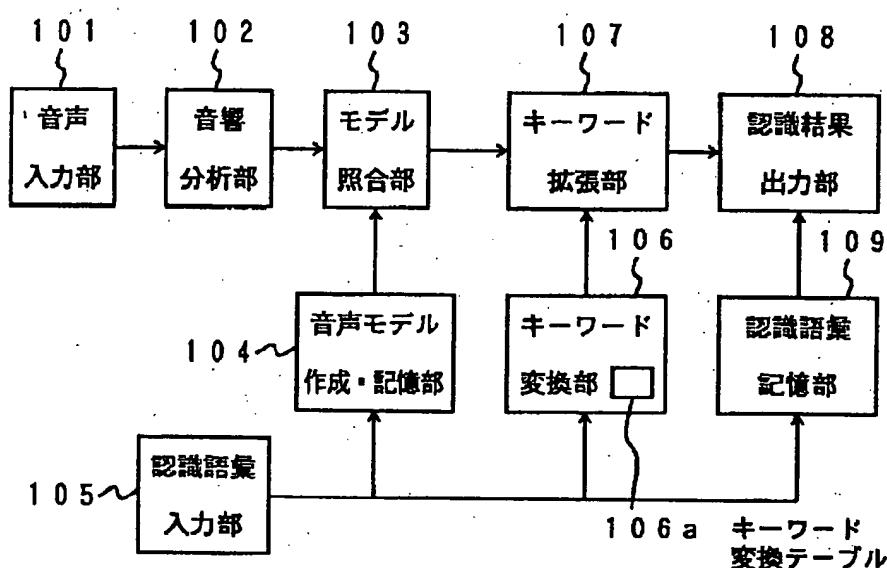
17

106…キーワード変換部
 107…キーワード拡張部
 108, 205, 308, 507, 606…認識結果出力部
 109, 407…認識語彙記憶部
 204…音声モデル記憶部

18

206…制約条件記憶部
 306, 506…別称登録部
 307…登録語彙類似性判定部
 309…警告出力部
 406…キーワード別認識結果出力部
 607…音声出力部

【図1】



【図7】

キーワード	類似度
田中	110
佐藤	130
加藤	90
ホテル	120
酒店	110
ガソリンスタンド	80

【図2】

キーワード変換テーブル

佐藤 → 加藤	～106a
加藤 → 佐藤	
商店 → 書店	
書店 → 商店	

【図4】

音節列	類似度
「あ」「さ」「ひ」「の」「あ」	86
「う」「さ」「ひ」「の」「あ」	92

(a)

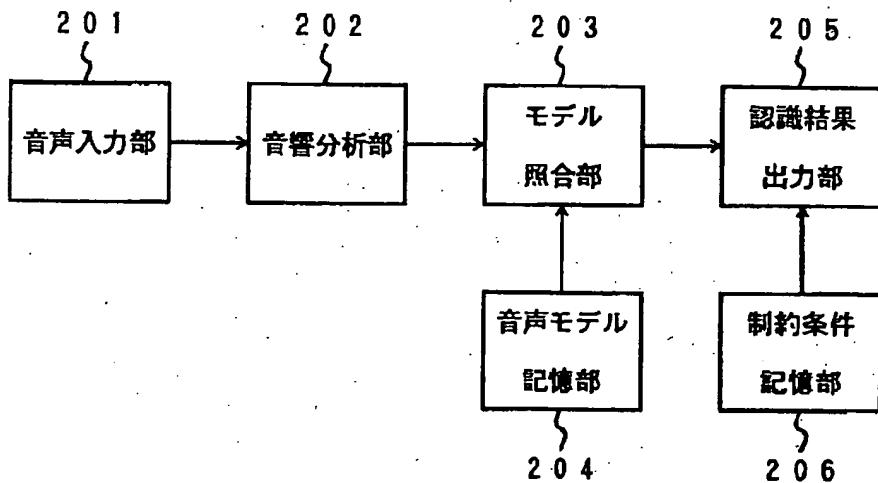
本実施形態による認識結果表示	
ホテル	酒店
佐藤ホテル 田中ホテル 加藤ホテル	佐藤酒店 田中酒店

従来の認識結果表示

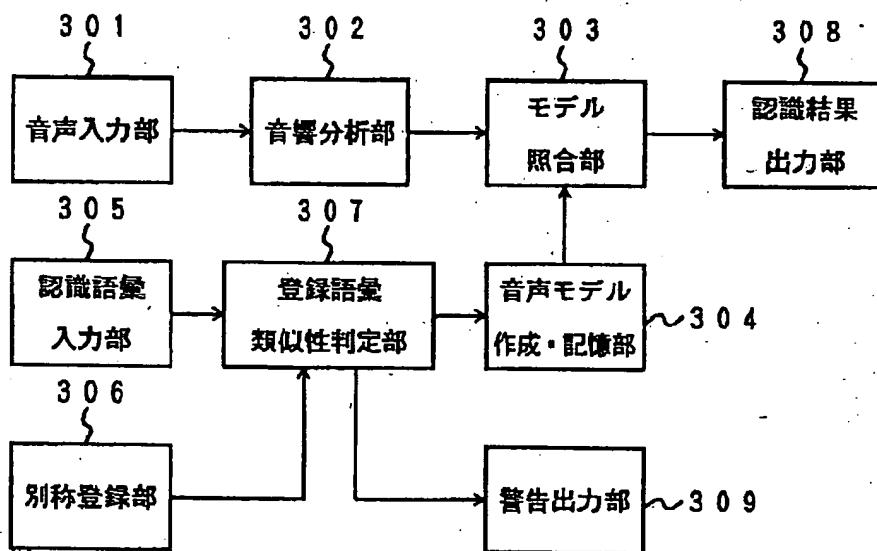
佐藤ホテル 佐藤酒店 田中ホテル 田中酒店 加藤ホテル 田中ガソリンスタンド
(b)

【図8】

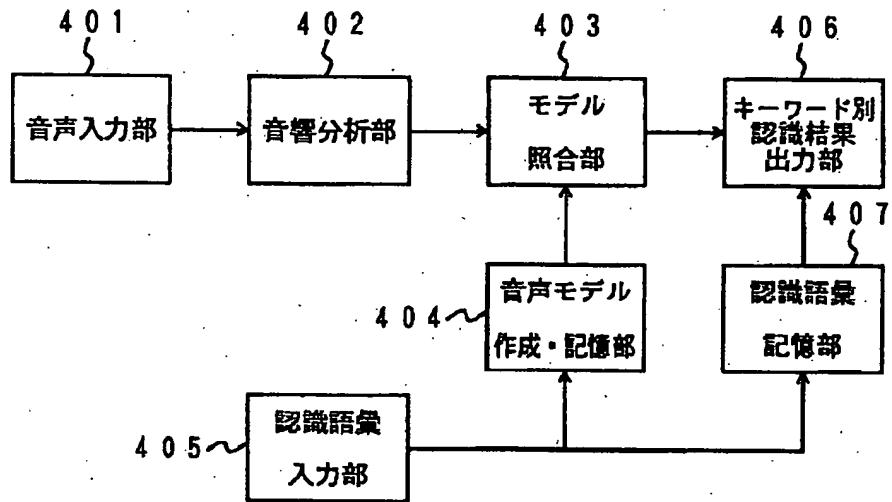
【 図3 】



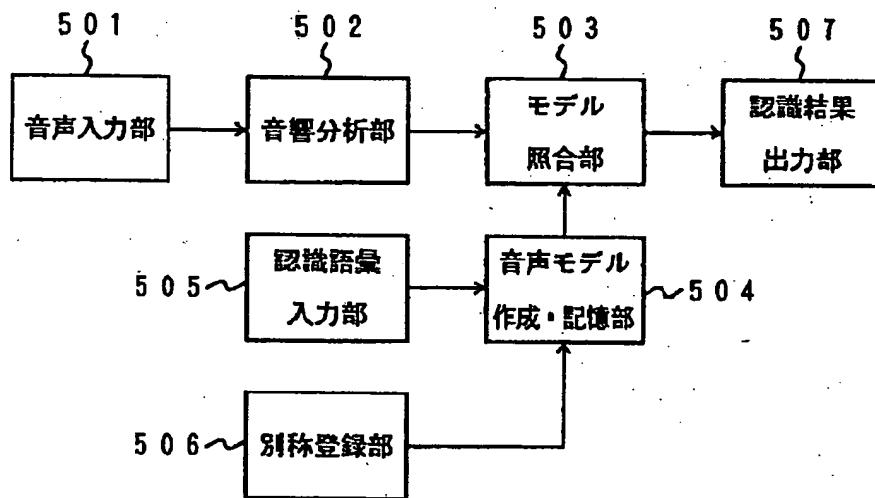
【 図5 】



【図6】



【図9】



【 図10 】

